

# A Survey on Heart Disease Prediction Techniques

G. Niranjana, Dr I. Elizabeth Shanthi

Department of Computer Science, Avinashilingam Institute for Home  
Science and Higher Education for Women, Coimbatore, Tamil Nadu, India

## ABSTRACT

Heart disease is the main reason for a huge number of deaths in the world over the last few decades and has evolved as the most life-threatening disease. The health care industry is found to be rich in information. So, there is a need to discover hidden patterns and trends in them. For this purpose, data mining techniques can be applied to extract the knowledge from the large sets of data. Many researchers, in recent times have been using several machine learning techniques for predicting the heart related diseases as it can predict the disease effectively. Even though a machine learning technique proves to be effective in assisting the decision makers, still there is a scope for developing an accurate and efficient system to diagnose and predict the heart diseases thereby helping doctors with ease of work. This paper presents a survey of various techniques used for predicting heart disease and reviews their performance.

**KEYWORDS:** Heart disease, data mining, machine learning, prediction

**How to cite this paper:** G. Niranjana | Dr I. Elizabeth Shanthi "A Survey on Heart Disease Prediction Techniques" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-2, February 2021, pp.68-71, URL: [www.ijtsrd.com/papers/ijtsrd38349.pdf](http://www.ijtsrd.com/papers/ijtsrd38349.pdf)



IJTSRD38349

Copyright © 2021 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



## 1. INTRODUCTION

Data mining is the process of examining large databases in order to discover the hidden patterns and correlations using statistics, machine learning, artificial intelligence and database technology. Tremendous amount of data is generated in medical field and it is important to mine those data for helping the practitioners in early diagnosis of disease. Heart disease causes immediate death and claims more lives each year than compared to all types of cancer or other major diseases. Heart disease prediction is very challenging area in the medical field because of several risk factors such as high blood pressure, high cholesterol, uncontrolled diabetes, abnormal pulse rate, obesity etc. Highly skilled and experienced physicians are required to diagnose the heart disease [1]. However, the death rate can be drastically reduced if the disease is detected at the early stages and also by adopting preventive measures. So, developing a heart disease prediction system is indispensable thereby preventing the patient's death through early diagnosis. The main objective of this paper is to do a survey on the previous research work in heart disease prediction and analyzing the techniques used.

The organization of the paper is as follows. Section 2 tells about the heart disease. Section 3 explains about the data mining algorithms for heart disease prediction. Section 4 deals with the literature review. Section 5 shows the observation and section 6 concludes the paper.

## 2. HEART DISEASE

Heart disease occurs when plaque develops in the arteries and blood vessels that lead to the heart. This plaque blocks

oxygen and important nutrients from reaching your heart. The types of heart disease are listed below:

1. Congenital heart disease
2. Arrhythmia
3. Coronary artery disease
4. Dilated cardiomyopathy
5. Myocardial infarction
6. Mitral valve prolapse
7. Pulmonary stenosis
8. Hypertrophic cardiomyopathy

## 3. DATA MINING TECHNIQUES FOR PREDICTION

### 3.1. DECISION TREE

Decision Tree is a type of Supervised Machine Learning algorithm, where it identifies several ways to split the data continuously based on certain parameter. The tree consists of two entities, namely decision node and leaf node. The decision nodes are where the data is split and the leaf node are the decisions or the final outcomes. Decision tree performs both classification and regression tasks. The tree model where the target variable takes a set of discrete values are called classification trees and when the target variable takes continuous values are called as regression trees. Decision tree is widely used in data mining, statistics and machine learning.

### 3.2. NAÏVE BAYES

A Naive Bayes classifier is a probabilistic machine learning algorithm based on Bayes theorem for binary and multi-class classification problems. A Naive Bayesian classifier is easy to build, with independent assumptions between predictors.

This assumption is most unlikely in real data. In Naive Bayes every pair of features being classified is independent of each other. It outperforms many other sophisticated classification models. Bayes' theorem is stated as follows  

$$P(A/B) = P(B/A).P(A)/P(B)$$

### 3.3. Random Forest

Random forest is a supervised learning algorithm which operates by constructing multiple decision trees for both classification as well as regression. The random forest algorithm creates decision trees based on data samples and then predicts the output for each of them and voting is done finally to select the best solution or output. The main advantage of using random forest is that it reduces the over-fitting and provides high accuracy.

### 3.4. SUPPORT VECTOR MACHINE

A Support vector machine is a discriminative classifier that finds a hyper plane in n-dimensional space to distinctly classify the data points. Given a labelled training data, the algorithm outputs an optimal separating hyper plane which divides the data or classes. Support vectors are the data points or class closer to the hyper plane. SVM is a supervised learning method used for both classification and regression tasks. It is widely used algorithm since it produces high accuracy with less computation power.

### 3.5. ARTIFICIAL NEURAL NETWORK

Neural network is a machine learning algorithm that is designed based on the human brain. Artificial neural network is an information processing technique that works like the way human brain process the information. ANN contains 3 layers input layer, hidden layer and the output layer that are interconnected by nodes which contains an activation function. Neural network is greatly used in data mining field for classifying large datasets. It enhances the data analysis technology.

## 4. LITERATURE REVIEW

In this section a detailed description of the previous work that has gone into the research related to the data mining technique for heart disease prediction are explained.

Purushottam et al. [2] (2015) designed a system that can efficiently discover the rules to predict the risk level of heart disease based on the given parameter. The rules are prioritized based on the user's requirements. The system uses the classification model by covering rules as C4.5Rules. WEKA tool is used for dataset analysis and Knowledge Extraction based on Evolutionary Learning (KEEL) tool to find out the classification decision rules. The dataset is taken from the Cleveland Clinic Foundation. It contains total 76 raw attributes, out of which only 14 of them are taken. This dataset contains several parameters like ECR, cholesterol, chest pain, fasting sugar, MHR and many more. The classification result of the decision tree is 87.4%.

AnchanaKhemphila et al. [3] (2011) proposed a classification approach using Multi-Layer Perceptron with Back Propagation learning algorithm and a feature selection algorithm to diagnose heart disease. Information Gain concept is used for feature selection. First the model uses the ANN with no information gain-based feature selection function; the accuracy in training dataset is 88.46% and 80.17% in the validation dataset. Further, the ANN is used

for classification after deducting the feature with lowest information gain. Now, the accuracy is 89.56% in training dataset and 80.99% in validation dataset. The result shows that feature selection helps increase computational efficiency while improving classification accuracy.

Nikhil Gawande et al. [4] (2017) proposed a system to classify heart disease using convolutional neural network. The system uses a CNN model to classify ECG signals. As it can be able to classify heart beat in different manner, there is no need for feature extraction. ECG signals have been taken from the MIT-BIH database. ECG signal is given as input to the system. Total 340 samples are trained and tested using CNN and the results are described in the confusion matrix. Even long ECG records can be classified in accurate manner and the distinct patient's records can be treated once the CNN is trained. The accuracy obtained is 99.46%.

V Krishnaiah et al. [5] (2014) introduces a fuzzy classification technique to diagnose the heart disease. The main objective of the research work is to predict the heart disease patients with more accuracy and to remove the uncertainty in unstructured data. The data source is collected from Cleveland Heart Disease database and Stalog Heart disease database. Cleveland database consist of 303 records and Stalog database consists of 270 records and remaining records are collected from different hospitals in Hyderabad.

M.A. Jabbar et al. [6] (2016) proposed a novel classification model Hidden Naïve Bayes classifier for heart disease prediction. The heart dataset is downloaded from the UCI repository. Heart stalog dataset contains 14 attributes and 270 instances. Hidden Naïve Bayes evaluation is performed using 10-fold cross validation. WEKA tool is used for hidden naïve bayes classification. The proposed approach performs pre-processing using discretization and IQR filters to improve the efficiency of Hidden Naïve Bayes. The performance results show that the HNB classifier model recorded 100% accuracy compared with NB classification model.

Jayshril S. Sonawane et al. [7] (2014) proposed a system to predict heart disease using multilayer perceptron architecture of neural network. The dataset is taken from the Cleveland heart disease database. The dataset containing 13 clinical attributes fed as input to the neural network and back propagation algorithm is used to train the data. Compared to the decision support system for predicting heart disease using multilayer perceptron and back propagation algorithm, the proposed system achieves highest accuracy of 98.58% for 20 neurons.

TanmayKasbe et al. [8] (2017) proposed fuzzy expert system for heart disease diagnosis. The database has been taken from the UCI repository and this dataset consist of 4 databases that are taken from V.A. Medical center, Long Beach, Cleveland clinic foundation, Hungarian institute of cardiology, Budapest and University hospital, Zurich, Switzerland. A total of 76 input attributes and 1 output attribute are in the dataset, out of this the proposed system uses 10 important input attributes and 1 output attribute. MATLAB software is used for developing fuzzy system. The fuzzy system consists of 3 steps i.e. fuzzification, rule base and defuzzification. The proposed fuzzy expert system

achieves 93.33% accuracy and better performance compared to previous work on same domain.

Aakashchauhan et al. [9] (2018) proposed a new rule to predict the coronary heart disease using evolutionary rule learning. Computational intelligence is used to discover the relationship between disease and patient. The dataset is taken from the Cleveland Heart Disease database. KEEL is a java-based tool used for the simulation of evolutionary learning. Classification model was developed by association rule and the results are evaluated. The system will help doctor to explore their data and predicts the coronary disease accurately.

Senthilkumarmohan et al. [10] (2019) proposed a novel method to improve the prediction accuracy of cardiovascular disease using hybrid machine learning techniques. The prediction model is designed with different combinations of features and several known machine learning classification techniques. Cleveland Heart Disease dataset is taken from the UCI repository. After feature selection 13 attributes are considered for further classification. R studio rattle is used for performing classification and the performance are evaluated. The prediction model uses hybrid random forest with a linear model and recorded 88.7% accuracy.

Sinkonnayak et al. [11] (2019) proposed a method to predict the heart disease by mining frequent items and classification techniques. The dataset is taken from UCI repository and pre-processing is done. The frequent item mining is used for filtering the attributes and then the variant classification techniques like Decision tree, Naïve Bayes, Support Vector Machine and KNN classification methods are used for predicting the heart disease at an early stage. R analytical tool is used for implementation. Out of these diverse data mining techniques Naïve Bayes achieves 88.67% accuracy in predicting the heart disease with attribute filtration. The performance is evaluated using ROC curve.

RahmaAtallah et al. [12] (2019) proposed a majority voting ensemble method to predict the heart disease in humans. Here the model classifies the patient based on the majority vote of diverse machine learning models in order to provide more accurate results. The dataset is taken from UCI repository and 14 predominant attributes are considered. In pre-processing Min-Max normalization is done. In order to analyze the data, a correlation value was calculated between each attribute and the target variable. It can be noted that the highest correlated features with the target attribute were Cp, Thalach, Oldpeak and Exang. For testing 4 types of classifier models are used namely Stochastic Gradient Descent (88%), KNN (87%), Logistic regression (87%) and Random forest (87%). These 4 models are combined in an ensemble model where the classification is done based on hard voting and finally the model achieves 90% accuracy.

Haritajagad and Jehankandawalla et al, [13] (2015) proposed a model to detect coronary artery disease using different data mining algorithms namely Decision Tree, Naïve Bayes and Neural Network. The parameters like patient age, sex, blood pressure etc. is taken during check up to evaluate the performance of these algorithms. Naïve Bayes proved to be the fastest among three algorithms. Decision tree algorithm reliability depends on the input data and it is difficult to deal with large datasets. Neural network is basically used when

dataset is small in size. There is no clear information about the accuracy level.

M. AkhilJabbar et al, [14] (2013) proposed a neural network model to classify the heart disease using artificial neural network and feature subset selection. The feature subset selection is a method that is used to reduce the dimensionality of the input data. By reducing the number of attributes, the number of diagnosis tests which are needed by doctors from patient are also reduced. The dataset is taken from Andhra Pradesh hospital and results show that accuracy is enhanced over the out-dated classification techniques. The results also show that this system is faster and precise.

Muhammad saqlain and wahidhussain et al, [15] (2016) proposed a multi nominal Naïve Bayes algorithm to detect the heart failure. The data are collected from Armed Forces Institute of Cardiology (AFIC), Pakistan in the form of medical records. It uses 30 variables. The proposed algorithm is compared with different classification algorithms like Logistic Regression, Neural Network, SVM, Random Forest and Decision Tree. The performance of Navies Bayes algorithm is measured in terms of Precision, Accuracy, Recall and Area Under the Curve (AUC). Naïve Bayes achieved highest accuracy of 86.7% and Area under the Curve (AUC) is 92.4% respectively.

## 5. OBSERVATION

From the literature review, it is observed that the dataset was taken from UCI repository containing Cleveland Heart Disease database and Stalog Heart disease database. The Naïve Bayes, Decision tree, Random forest and neural network algorithms are predominantly used data mining techniques for predicting the heart disease with highest accuracy. But the authors should concentrate on minimizing the time utilization. Most of the research works have used existing machine learning techniques or combined two or more existing techniques to improve the prediction performance. Rather than taking vote or combining the existing algorithms the researchers should propose new algorithms for prediction. The researchers should explore more deep learning concepts in order to improve the prediction performance. Since the heart disease is very dangerous a fast and reliable system need to be developed.

## 6. CONCLUSION

The main motive for conducting this survey is to comprehend the work of different authors and also to analyze how accurately we can predict the heart disease. These authors have proposed different data mining algorithms but there are certain limitations like neural networks performs well only with structured dataset [16]. MATLAB, Python and WEKA tool are widely used technologies for implementing these algorithms. However different algorithms generate different accuracy that purely depends on the size of the dataset, number of attributes selected [17] and tools used for implementation. The future works should concentrate on proposing new algorithms to achieve better performance and to minimize the time utilization.

## REFERENCES

- [1] M. A. Jabbar et.al" Prediction of heart disease using Random forest and Feature subset selection ",AISC SPRINGER, vol 424,pp187-196(2015)



- [2] Purushottam, K. Saxena and R. Sharma, "Efficient heart disease prediction system using decision tree," International Conference on Computing, Communication & Automation, Noida, 2015, pp. 72-77.
- [3] Khemphila and V. Boonjing, "Heart Disease Classification Using Neural Network and Feature Selection," 2011 21st International Conference on Systems Engineering, Las Vegas, NV, 2011, pp. 406-409.
- [4] N. Gawande and A. Barhatte, "Heart diseases classification using convolutional neural network," 2017 2nd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2017, pp. 17-20.
- [5] V. Krishnaiah, M. Srinivas, G. Narsimha and N. S. Chandra, "Diagnosis of heart disease patients using fuzzy classification technique," International Conference on Computing and Communication Technologies, Hyderabad, 2014, pp. 1-7.
- [6] M. A. Jabbar and S. Samreen, "Heart disease prediction system based on hidden naïve bayes classifier," 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, 2016, pp. 1-5.
- [7] J. S. Sonawane and D. R. Patil, "Prediction of heart disease using multilayer perceptron neural network," International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, 2014, pp. 1-6.
- [8] T. Kasbe and R. S. Pippal, "Design of heart disease diagnosis system using fuzzy logic," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, 2017, pp. 3183-3187.
- [9] A. Chauhan, A. Jain, P. Sharma and V. Deep, "Heart Disease Prediction using Evolutionary Rule Learning," 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, 2018, pp. 1-4.
- [10] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019.
- [11] S. Nayak, M. K. Gourisaria, M. Pandey and S. S. Rautaray, "Prediction of Heart Disease by Mining Frequent Items and Classification Techniques," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 607-611, doi: 10.1109/ICCS45141.2019.9065805.
- [12] R. Atallah and A. Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method," 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923053.
- [13] Harita Jagad and Jehan Kandawalla, "Detection of Coronary Artery Diseases using Data Mining Techniques," International Journal on Recent And Innovation Trends in Computing And Communication ISSN:2321-8169 Vol.3 issue 11, pp.60906092, 2015.
- [14] Jabbar, M. Akhil, B. L. Deekshatulu, and Priti Chandra. "Classification of heart disease using artificial neural network and feature subset selection." Global Journal of Computer Science and Technology Neural & Artificial Intelligence 13, no. 3 (2013).
- [15] Muhammad Saqlain, Wahid Hussain, Nazar A. Saqib et al, "identification of heart failure by using unstructured data of cardiac patients", 45th International Conference on Parallel Processing Workshops, DOI 10.1109/ICPPW.2016.66, pp.426431, 2016.
- [16] B. Ganeswar and M. R. E. Jebarani, "A review on prediction and diagnosis of heart failure," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIECS), Coimbatore, 2017, pp. 1-3.
- [17] A. L. Rane, "A survey on Intelligent Data Mining Techniques used in Heart Disease Prediction," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 208-213.